THE USE OF QUADRATIC PROGRAMMING IN STOCHASTIC LINEAR PROGRAMMING

E. M. L. Beale*
Consultant, The RAND Corporation

P-2404

15

August 15, 1961

*(CEIR(UK) Limited)
London, England

Paper to be presented at the Eighth Annual International Meeting of TIMS, Brussels, Belgium, August 23-26, 1961.

THE USE OF QUADRATIC PROGRAMMING IN STOCHASTIC LINEAR PROGRAMMING

E. M. L. Beale*

Consultant, The RAND Corporation

P-2404

15

August 15, 1961

*(CEIR(UK) Limited)
London, England

Paper to be presented at the Eighth Annual International Meeting of TIMS, Brussels, Belgium, August 23-26, 1961.

SUMMARY

This paper presents a simple method of allowing for uncertainties in the constant terms (i.e. right hand sides) of a linear programming problem, and hence producing realistic "safety margins" in the solution. This is done by fitting a mixture of uniform distributions to the assumed distributions of these right hand sides, and using a particular quadratic programming algorithm. ( )

THE USE OF QUADRATIC PROGRAMMING IN STOCHASTIC LINEAR PROGRAMMING

1. INTRODUCTION

It is well known that linear programming has an uncompromising addiction to basic solutions. This means that it will urge us to "go all out" for our objective, and cannot be used to estimate the "safety margins" that often have to be provided in practical situations to guard against uncertainties of one sort or another. As Dantzig (1956) pointed out, stochastic linear programming is the theoretical answer to this problem, but it does not seem to have been applied extensively. This is probably because the computations involved are usually rather heavy, and the available data on chance effects are rarely precise enough to support a heavy computation.

The purpose of this paper is to show that an important special class of such problems can be solved fairly easily using a standard quadratic programming algorithm. As far as I know, this algorithm has not yet been programmed for a computer, but I hope that it will be in the not too distant future. It should then be worth while to feed in estimates of the uncertainties, and their consequences, into the computations, even if these estimates are quite crude.

2. THE PROBLEM

The class of problems considered here has been discussed by many authors: Dantzig (1955), Beale (1955), Ferguson and Dantzig (1956), Elmaghraby (1959), (1960), Dantzig and Madansky

(1960), and Madansky (1960).

The nonstochastic version of the model contains equations of the form

$$(1) \qquad g_i'x = b_i,$$

where $g_i'x$ is some linear function of the variables of the problem, denoted collectively by the vector x, and where b_i is a number that one pretends is known exactly when it really is not. For example, it may be the volume of sales in some future time-period. The problem is to choose a nonnegative x satisfying all these equations so as to minimize some linear function $c'x$.

In the stochastic version of the model, the b_i are regarded as random variables with known probability distributions. The vector x has to be chosen before we know the actual values of the b_i; but we are no longer required to satisfy the equations (1) exactly. Instead we have a penalty of $f_i^-(\geq 0)$ for each unit by which $g_i'x$ falls short of b_i, and a penalty of $f_i^+(\geq 0)$ for each unit by which $g_i'x$ exceeds b_i, to be added to the direct cost $c'x$. This penalty may represent a tangible loss, such as failure to sell one's goods because the demand is inadequate, or an intangible loss of goodwill caused by failure to meet a requirement, or a combination of the two. The problem is then to choose a nonnegative x that minimizes the expected total cost

$$(2) \qquad T = c'x + \Sigma P_i,$$

where P_i denotes the mean value of the penalty incurred on the i-th equation. To represent P_i mathematically, we write (1) in the form

$$(3) \qquad \underset{\sim}{g}_i'\underset{\sim}{x} - y_i = b_i,$$

where

$$(4) \qquad y_i = y_i^+ - y_i^-, \; y_i^- \geq 0, \; y_i^+ \geq 0,$$

and

$$(5) \qquad P_i = E(f_i^- \, y_i^- + f_i^+ \, y_i^+).$$

In practice there will often be no penalty cost if $y_i > 0$, so that $f_i^+ = 0$. The nonstochastic model would then normally appear in the form $\underset{\sim}{g}_i'\underset{\sim}{x} \geq b_i$.

In fact it is convenient to reduce the general problem to one with $f_i^+ = 0$, by adding f_i^+ times equation (3) to equation (2). We then have

$$(6) \qquad T = -\sum_i f_i^+ \, \bar{b}_i + (\underset{\sim}{c} + \sum_i f_i^+ \, \underset{\sim}{g}_i)'\underset{\sim}{x} + \sum_i P_i'$$

where $\bar{b}_i$ denotes the mean value of b_i, and

$$(7) \qquad P_i' = E((f_i^+ + f_i^-)y_i^-).$$

It is perhaps worth noting that, although we have assumed that f_i^+ and f_i^- are both nonnegative, we really only require that $f_i^+ + f_i^-$ be nonnegative to justify the mathematics.

Given a problem with $f_i^+ = 0$, it is convenient to write

$$(8) \qquad \bar{y}_i = \bar{b}_i - g_i' x,$$

and to refer to $\bar{y}_i$ as the safety margin, or margin, in the i-th constraint. The essential contribution of the stochastic feature of the model is to provide a means of estimating proper values for these margins. In the nonstochastic model they are zero on all the operative constraints.

Another general point to be noted is that the correlations between different b_i do not enter the problem. They have no effect on the mean value of the penalty cost — although of course they could have an appreciable effect on its variance.

3. EXISTING METHODS OF SOLUTION

Methods for solving these problems are known. If all the probability distributions are discrete, any 2-stage linear programming problem can be reduced to a vast linear programming problem, which can nevertheless be solved using the decomposition principle of Dantzig and Wolfe (1960), as Dantzig and Madansky (1960) have pointed out. The simple type of penalty cost considered here is easier than this. It leads immediately to a convex separable nonlinear objective function, which can be handled by the methods of Charnes and Lemke (1954).

Elmaghraby (1960) has suggested that a continuous distribution function could be handled using the Lagrangian differential gradient method of Hurwicz (1957). But the limited computational experience with these methods is not very encouraging.

4. THE PROPOSED METHOD OF SOLUTION

The starting point for the approach proposed here is the fact, noted by Dantzig (1955), that if the random variable b_1 has a uniform (rectangular) distribution over some specified range then the objective function is quadratic. This follows from the fact that (assuming $f_1^+ = 0$), $dP_1/d\bar{y}_1$ equals minus f_1^- times the probability that $y_1 < 0$. In fact we find that if b_1 is uniformly distributed between $\bar{b}_1 - r_1$ and $\bar{b}_1 + r_1$, then

$$P_1 = 0 \qquad \text{if} \qquad \bar{y}_1 \geq r_1,$$

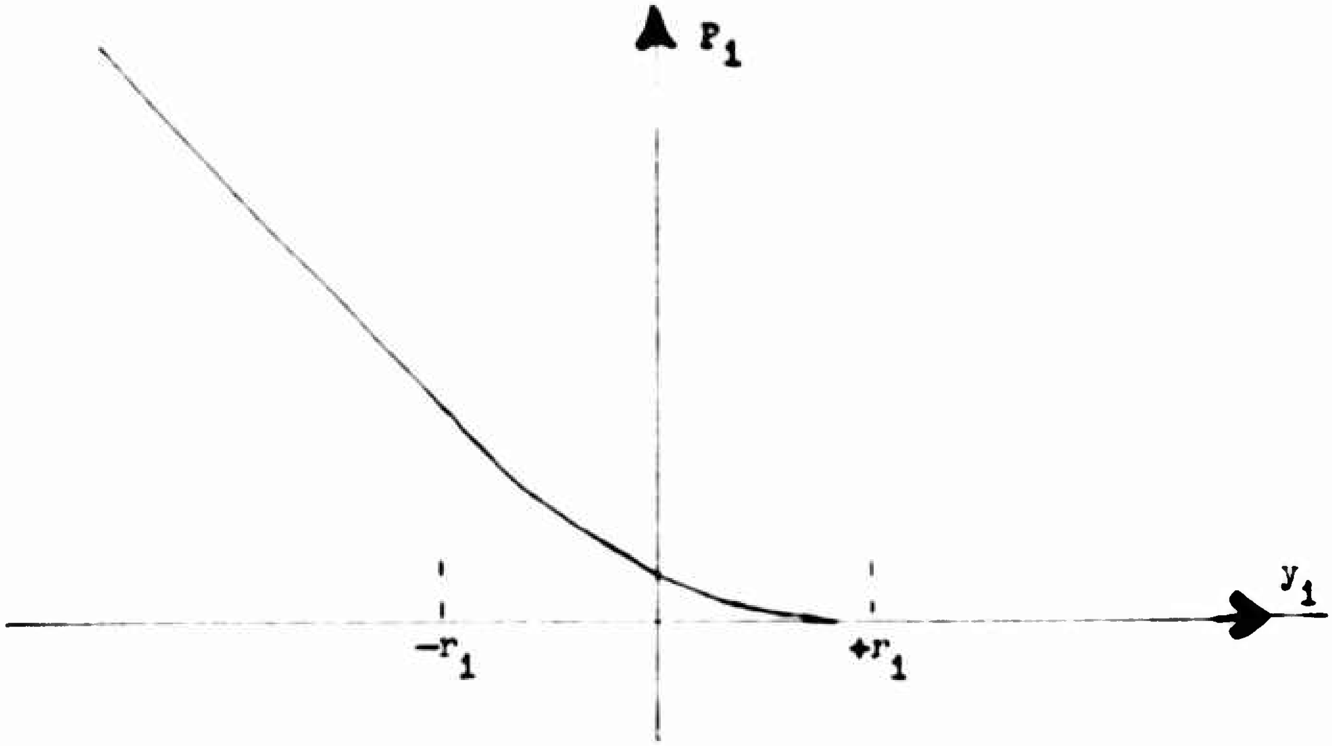
$$P_1 = \int_0^{(r_1 - \bar{y}_1)} f_1^- t \frac{1}{2r_1} dt$$

$$= \frac{f_1^-}{4r_1} (r_1 - \bar{y}_1)^2 \qquad \text{if} \qquad -r_1 < \bar{y}_1 < r_1,$$

$$P_1 = \int_{-r_1 - \bar{y}_1}^{r_1 - \bar{y}_1} f_1^- t \frac{1}{2r_1} dt$$

$$= -f_1^- \bar{y}_1 \qquad \text{if} \qquad \bar{y}_1 \leq -r_1.$$

This function, plotted against $\bar{y}_1$, looks as follows.



And it turns out that this function can be represented very simply, by writing

(9) $$\bar{y}_1 = r_1 + x_{11} - x_{21} - x_{31},$$

$$x_{11} \geq 0, \; x_{21} \geq 0, \; x_{31} \geq 0,$$

(10) $$P_1 = r_1^- (x_{21} + x_{31}^2/4r_1).$$

In other words the constraints (1) or (3) are replaced by

(11) $$\underset{\sim}{\xi}_1' \underset{\sim}{x} - x_{11} + x_{21} + x_{31} = \bar{b}_1 + r_1,$$

and (2) is replaced by

$$(12) \qquad T = \underset{\sim}{c}'\underset{\sim}{x} + \sum_i f_i^-(x_{2i} + x_{3i}^2/4r_i).$$

To verify these formulae, first consider the situation when $\bar{y}_i > r_i$. Then it is clear that we can put $x_{2i} = x_{3i} = 0$, and $P_i = 0$, as it should be. Now suppose that $\bar{y}_i < r_i$. Then we cannot put $x_{2i} = x_{3i} = 0$, and P_i will have to be greater than zero. But we would be crazy not to put $x_{1i} = 0$, since a positive x_{1i} would involve an unnecessarily large x_{2i} or x_{3i} at some extra expense and no benefit. So we put $x_{1i} = 0$ and make up the difference between $\bar{y}_i$ and r_i by $x_{2i} + x_{3i}$. Now $\partial P_i/\partial x_{2i} = f_i^-$, and $\partial P_i/\partial x_{3i} = f_i^- x_{3i}/2r_i$. So, if we have to increase $x_{2i} + x_{3i}$, it is cheaper to increase x_{3i} if $x_{3i} < 2r_i$, and otherwise it is cheaper to increase x_{2i}. Hence, as $\bar{y}_i$ decreases from r_i, P_i increases quadratically to r_i when $\bar{y}_i = -r_i$, and thereafter P_i increases linearly. In fact P_i is represented faithfully in all cases.

If it is more convenient, we can handle a problem with $f_i^+ \neq 0$ directly, by writing

$$(13) \quad P_i = f_i^+ x_{1i} + f_i^- x_{2i} - f_i^+ x_{3i} + (f_i^+ + f_i^-)x_{3i}^2/4r_i.$$

The problem of minimizing T subject to these constraints is a typical quadratic programming problem in which the number of quadratic terms equals the number of constraints with random right hand sides, which is almost certainly much smaller than the total number of variables in the problem. The problem is therefore suited to the algorithm outlined on pp. 235–236 of Beale (1959). This is essentially the algorithm presented by Beale (1955) and more fully in Beale (1959). The only difference is that the objective function is stored, not as a vast square matrix, but as a set of (r+1) linear forms λ_0, λ_1, ..., λ_r, it being understood that

$$T = \lambda_0 + \frac{1}{2}\lambda_1^2 + \dots + \frac{1}{2}\lambda_r^2.$$

So we can deal with uniformly distributed right hand sides fairly easily. In some situations we may know so little about the true distributions of these right hand sides that such an assumption is as good as any. But in other situations this will not be very satisfactory. For example, suppose we have a production scheduling problem in which sales are treated as random variables, but we wish to lay down the production schedule in advance. Then the nonstochastic model will contain some constraints of the form

$$\begin{aligned} I_0 + R_1 &\geq S_1, \\ I_0 + R_1 + R_2 &\geq S_1 + S_2 \\ &\cdots\cdots \end{aligned}$$

representing the facts that initial inventory plus total production up to the end of any time-period must not be less than total sales up to the end of the time-period. Now we might possibly assume that S_1 has a uniform distribution. But it would hardly be consistent to assume that $S_1 + S_2$ also had a uniform distribution.

A possibly more fundamental objection to the uniform distribution is that it assumes that there is no danger that the random variable will differ from its mean by more than 1.732 standard deviations. And yet it is clear that in some situations wider safety margins are needed.

But now we observe that the analysis can be carried through if the distribution of b_i is a mixture of a small number of uniform distributions almost as easily as if it is a single uniform distribution. Suppose that there is a probability p_{ij} that b_i is uniformly distributed between $m_{ij} - r_{ij}$ and $m_{ij} + r_{ij}$, for $j = 1, \ldots, k_i$. Then if we replace the i-th constraint by the k_i constraints

$$(14) \qquad \underset{\sim}{g}_i'\underset{\sim}{x} - x_{1ij} + x_{2ij} + x_{3ij} = m_{ij} + r_{ij},$$

we have

$$(15) \qquad T = \underset{\sim}{c}'\underset{\sim}{x} + \sum_i \sum_j f_i^{-} p_{ij} (x_{2ij} + x_{3ij}^2/4r_{ij}).$$

This approach would be unattractive if we had to use a large value of k_i, i.e. a large number of uniformly distributed components, for each constraint with a random right hand side. One might think, for example that several components would be needed to approximate a normal frequency function by a step function. But the penalty cost is not represented by the frequency function, but by the integral of the cumulative distribution function. And the process of integrating twice smooths out the corners in the approximation in a most gratifying way. An approximation (fitted by the method of moments) to the normal distribution by just two uniform distributions gives an expected penalty cost that is correct to within 15% for any margin less than 1.5 standard deviations. For a margin between 1.5 and 2.5 standard deviations the approximation exaggerates the expected penalty cost by up to 50%. This may seem an undesirably large exaggeration, but it would arise equally if one used a genuine normal distribution with a standard deviation overestimated by 5%. So in practice it will often be unrealistic to try for greater precision than that given by two uniformly distributed components. The details of the method of fitting are discussed in the appendices, the results for a normal distribution being illustrated in Figs. 1 and 2.

We therefore need just one extra constraint, and two quadratic variables in the objective function, for each constraint in the original problem with a random right hand side. Since this will usually apply to not more than half the constraints, this is not a large addition.

5. THE RELATIONSHIP WITH CHANCE-CONSTRAINED LINEAR PROGRAMMING

It is of interest to compare this "2-stage" approach to the problem of unknown right hand sides with chance-constrained programming, introduced by Charnes, Cooper and Symonds (1958). Of course this problem is not the most general possible problem for either chance-constrained or 2-stage linear programming;

but it is often illuminating to compare different approaches with reference to a specific simple class of problems.

One difference is that the 2-stage program minimizes the total cost, consisting of a direct cost and a penalty cost representing the average level of failure to satisfy the original inequality constraints. In chance-constrained programming one pre-selects a tolerable long-term failure level, and simply minimizes the direct cost without exceeding this failure level on the average. Such an approach may be advantageous if the penalty costs are hard to assess quantitatively. For our problem it can easily be incorporated into the 2-stage model by using parametric programming. As Beale (1959) indicates, parametric quadratic programming is straightforward as long as the parameter is confined to the right hand sides and the linear part of the objective function. To apply it in this context we must therefore keep the penalty cost function fixed, and gradually scale down the direct cost to represent a gradually increasing relative importance to be assigned to failure. Corresponding to each parameter value, one can compute the direct cost and the average level of infeasibility, measured by the average penalty cost incurred. One can then select that parameter value corresponding to the selected tolerance for infeasibility.

But the more fundamental difference between 2-stage and chance-constrained programming is that, for any given values of the variables and the random elements, the infeasibility is measured in chance-constrained programming essentially by the number of constraints that are violated, and not by the extent to which they are violated. (Actually this is an oversimplification. Violations of some constraints may have to be regarded as more serious than violations of others, to allow the overall probabilities of violating different constraints to balance in an optimum program.) The question of which of these is the more reasonable measure must depend on the application. Madansky (1960) suggests an alternative form of chance-constrained programming in which the infeasibility is measured by whether or not all the constraints are satisfied simultaneously. To apply this to our problem, we would need to know the joint distribution of all the random variables, whereas the marginal distributions suffice for the other approaches.

It is possible to compromise between our formulation of the problem and chance-constrained programming, by accepting our measure of infeasibility, computing the parametric family of optimum solutions for different relative weights on cost and feasibility, and then selecting from this family the solution with a specified average total number of violated constraints. In some circumstances this might be a very practical procedure, giving an easily comprehensible if indirect control over the level of infeasibility.

APPENDIX 1

APPROXIMATING A SYMMETRICAL DISTRIBUTION BY A MIXTURE OF UNIFORM DISTRIBUTIONS

To apply the methods of Section 4, the assumed distribution for the random variable b must be approximated by a mixture of a small number of uniform distributions. It is natural to use the method of moments to define this approximation. This is a somewhat arbitrary decision, but it seems natural, because

(a) the moments of the given distribution can usually be readily evaluated, and

(b) it tends to emphasize the tails of the distribution, where a good fit is most important.

The problem is greatly simplified if b has a symmetrical distribution. It is then clear that we should make the mean of each component equal to the mean of the distribution, and it only remains to choose the proportions p_i and half-ranges r_i to be allotted to the different components. With k components, we can fit up to the (4k − 2)-th moment, and the equations to be satisfied are as follows:

$$\begin{aligned}
\Sigma p_i \qquad &= 1 \\
\Sigma p_i \; r_i^2 \qquad &= 3\mu_2 \\
\Sigma p_i \; r_i^4 \qquad &= 5\mu_4 \\
&\cdots \\
\Sigma p_i \; r_i^{(4k-2)} \qquad &= (4k-1)\mu_{4k-2},
\end{aligned}$$

where μ_i denotes the i-th moment about the mean of the (symmetric) distribution being fitted.

Writing $r_i^2 = \lambda_i$, $(2i + 1)\mu_{2i} = \nu_i$, we find that our equations reduce to

$$\begin{aligned}
&\Sigma p_i &&= 1 \\
&\Sigma p_i \lambda_i &&= \nu_1 \\
&&&\cdots \\
&\Sigma p_i \lambda_i^{2k-1} &&= \nu_{2k-1}.
\end{aligned}$$

These are the standard formulae for the weights and ordinates for Gauss-type quadrature formulae. For convenience we record the solutions for k = 1, 2 or 3.

If $k = 1$, $\lambda_1 = \nu_1$, $p_1 = 1$.

If $k = 2$, λ_1 and λ_2 are the roots of the quadratic equation

$$\begin{vmatrix} x^2 & x & 1 \\ \nu_2 & \nu_1 & 1 \\ \nu_3 & \nu_2 & \nu_1 \end{vmatrix} = 0,$$

and

$$p_1 = (\nu_1 - \lambda_2)/(\lambda_1 - \lambda_2)$$
$$p_2 = (\nu_1 - \lambda_1)/(\lambda_2 - \lambda_1).$$

If $k = 3$, λ_1, λ_2 and λ_3 are the roots of the cubic equation

$$\begin{vmatrix} x^3 & x^2 & x & 1 \\ v_3 & v_2 & v_1 & 1 \\ v_4 & v_3 & v_2 & v_1 \\ v_5 & v_4 & v_3 & v_2 \end{vmatrix} = 0,$$

and

$$p_1 = (v_2 - v_1(\lambda_2 + \lambda_3) + \lambda_2\lambda_3)/(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)$$

$$p_2 = (v_2 - v_1(\lambda_1 + \lambda_3) + \lambda_1\lambda_3)/(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3)$$

$$p_3 = (v_2 - v_1(\lambda_1 + \lambda_3) + \lambda_1\lambda_2)/(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2).$$

For a normal distribution with variance σ^2, $v_1 = 1.3\sigma^2$, $v_2 = 1.3\sigma^4$, $v_3 = 1.3.5.7\sigma^6$, $v_4 = 1.3.5.7.9\sigma^8$, $v_5 = 1.3.5.7.9.11\sigma^{10}, \ldots$ and we have the following results.

k	p_i	v_i
1	1	1.7321σ
2	0.1838	2.8570σ
	0.8162	1.3556σ
3	0.0154	3.7504σ
	0.3446	2.3668σ
	0.6400	1.1544σ

The fitted frequency functions are plotted in Fig. 1, and the resulting average penalty costs, assuming $\sigma = 1$, $r_1^- = 1$, $r_1^+ = 0$, are plotted on a logarithmic scale in Fig. 2. It will be seen that the single uniform distribution ($k = 1$) gives a very adequate fit for margins less than σ, but underestimates the penalty for larger margins. The mixture of 2 uniform distributions ($k = 2$) gives an adequate fit for margins less than 2.5σ, and this should be satisfactory for most practical purposes. The improvement obtained by taking a third component ($k = 3$) is not spectacular. One gets an adequate fit right out to 3.5σ, but there is an awkward trough in the curve around 2.3σ where the penalty is underestimated by 34%. This is perhaps because the tenth moment, which is used to fit this mixture, gives too much weight to the extreme tails of the distribution. A better fit over a somewhat shorter range of margins could perhaps be obtained by fitting the moments of a truncated normal distribution.

APPENDIX 2

APPROXIMATING AN UNSYMMETRICAL DISTRIBUTION BY A MIXTURE OF UNIFORM DISTRIBUTIONS

The problem of fitting a mixture of uniform distributions by the method of moments to an unsymmetrical distribution seems to be much more awkward than fitting a symmetrical distribution. We now have to choose 3 parameters for each component: the proportion p_i, the mean m_i and the half-range r_i. With k components we can therefore fit up to the (3k - 1)-th moment. Without loss of generality we can assume that the mean of the distribution to be fitted is zero.

The r-th moment of the mixture is given by

$$\sum_i \frac{p_i}{2(r+1)r_i} \{(m_i + r_i)^{r+1} - (m_i - r_i)^{r+1}\}.$$

If μ_r denotes the r-th moment about the mean of the distribution to be fitted, we therefore obtain the following equations for fitting 2 components:

$$(A2.1)\begin{cases} p_1 + p_2 = 1 \\ p_1 m_1 + p_2 m_2 = 0 \\ p_1(m_1^2 + \frac{1}{3} r_1^2) + p_2(m_2^2 + \frac{1}{3} r_2^2) = \mu_2 \\ p_1(m_1^3 + m_1 r_1^2) + p_2(m_2^3 + m_2 r_2^2) = \mu_3 \\ p_1(m_1^4 + 2m_1^2 r_1^2 + \frac{1}{5} r_1^4) + p_2(m_2^4 + 2m_2^2 r_2^2 + \frac{1}{5} r_2^4) = \mu_4 \\ p_1(m_1^5 + \frac{10}{3} m_1^3 r_1^2 + m_1 r_1^4) + p_2(m_2^5 + \frac{10}{3} m_2^3 r_2^2 + m_2 r_2^4) = \mu_5 . \end{cases}$$

There does not seem to be any neat way of solving these equations. But we can tackle them by solving the first 2 equations for p_1 and p_2 in terms of m_1 and m_2, substituting these values into the 3-rd and 4-th equation and solving for r_1^2 and r_2^2 in terms of m_1 and m_2, and then substituting all these values into the 5-th and 6-th equations to determine m_1 and m_2. We find that

(A2.2) $$p_1 = -m_2/(m_1 - m_2), \qquad p_2 = m_1/(m_1 - m_2)$$

(A2.3) $$r_1^2 = (-\frac{\mu_3}{m_2} + 3\mu_2 + 2m_1m_2 - m_1^2), \quad r_2^2 = (-\frac{\mu_3}{m_1} + 3\mu_2 + 2m_1m_2 - m_2^2).$$

When we substitute these values into the last 2 equations of (A2.1) it is convenient to write

(A2.4) $$\lambda_1 = m_1 + m_2, \quad \lambda_2 = m_1m_2.$$

The equations then reduce to

(A2.5) $$A_{10} + A_{11}\lambda_1 + A_{12}\lambda_1^2 = 0$$

(A2.6) $$A_{20} + A_{21}\lambda_1 + A_{22}\lambda_1^2 + A_{23}\lambda_1^3 = 0,$$

where

$$A_{10} = -\mu_3^2 + (9\mu_2^2 - 5\mu_4)\lambda_2 - 12\mu_2\lambda_2^2 - 16\lambda_2^3$$

$$A_{11} = +8\mu_3\lambda_2$$

$$A_{12} = +4\lambda_2^2,$$

$$A_{20} = (18\mu_2\mu_3 - 3\mu_5)\lambda_2 + 8\mu_3\lambda_2^2$$

$$A_{21} = -3\mu_3^2 - 12\mu_2\lambda_2^2 - 16\lambda_2^3$$

$$A_{22} = +4\mu_3\lambda_2$$

$$A_{23} = +4\lambda_2^2.$$

We can manipulate these equations to solve for λ_1 in terms of λ_2, and obtain the formula

$$\text{(A2.7)} \quad \lambda_1 = \frac{\mu_3^3 + (5\mu_3\mu_4 - 9\mu_2^2\mu_3)\lambda_2 + (3\mu_5 - 6\mu_2\mu_3)\lambda_2^2 + 8\mu_3\lambda_2^3}{\lambda_2(6\mu_3^2 + (5\mu_4 - 9\mu_2^2)\lambda_2)}.$$

Finally, substituting this back into (A2.5) we obtain a sextic equation for λ_2, which can be written as

$$\text{(A2.8)} \quad a_0 + a_1\lambda_2 + a_2\lambda_2^2 + a_3\lambda_2^3 + a_4\lambda_2^4 + a_5\lambda_2^5 + a_6\lambda_2^6 = 0,$$

where

$$a_0 = 16\mu_3^6,$$

$$a_1 = 80\mu_3^4\mu_4 - 144\mu_2^2\mu_3^4,$$

$$a_2 = 168\mu_3^3\mu_5 - 25\mu_3^2\mu_4^2 + 90\mu_2^2\mu_3^2\mu_4 - 768\mu_2\mu_3^4 - 81\mu_2^4\mu_3^2,$$

$$a_3 = 240\mu_3\mu_4\mu_5 - 432\mu_2^2\mu_3\mu_5 - 125\mu_4^3 + 675\mu_2^2\mu_4^2 - 1200\mu_2\mu_3^2\mu_4$$
$$- 1215\mu_2^4\mu_4 - 128\mu_3^4 + 2160\mu_2^3\mu_3^2 + 729\mu_2^6,$$

$$a_4 = 36\mu_5^2 - 144\mu_2\mu_3\mu_5 - 300\mu_2\mu_4^2 - 320\mu_3^2\mu_4 + 1080\mu_2^2\mu_4$$
$$+ 720\mu_2^2\mu_3^2 - 972\mu_2^5,$$

$$a_5 = 192\mu_3\mu_5 - 400\mu_4^2 + 1440\mu_2^2\mu_4 - 384\mu_2\mu_3^2 - 1296\mu_2^4,$$
$$a_6 = 256\mu_3^2.$$

The suggested computational procedure is therefore as follows:

1. Compute the coefficients of (A2.8), and find a negative real root of the equation. If there are no negative roots, then the problem is insoluble. If there are any, then there must be an even number. It seems likely that the smallest in absolute value will be appropriate, but this is not exactly clear.

2. Given the value of λ_2, deduce λ_1 from (A2.7).

3. Find the roots of the quadratic equation

$$x^2 - \lambda_1 x + \lambda_2 = 0.$$

These are m_1 and m_2.

4. Deduce p_1, p_2, r_1^2 and r_2^2 from (A2.2) and (A2.3). Since $\lambda_2 < 0$, m_1 and m_2 must have opposite signs, so p_1 and p_2 are certain to be positive. But r_1^2 or r_2^2 might prove to be negative, indicating that the problem has no solution, at least for this value of λ_2.

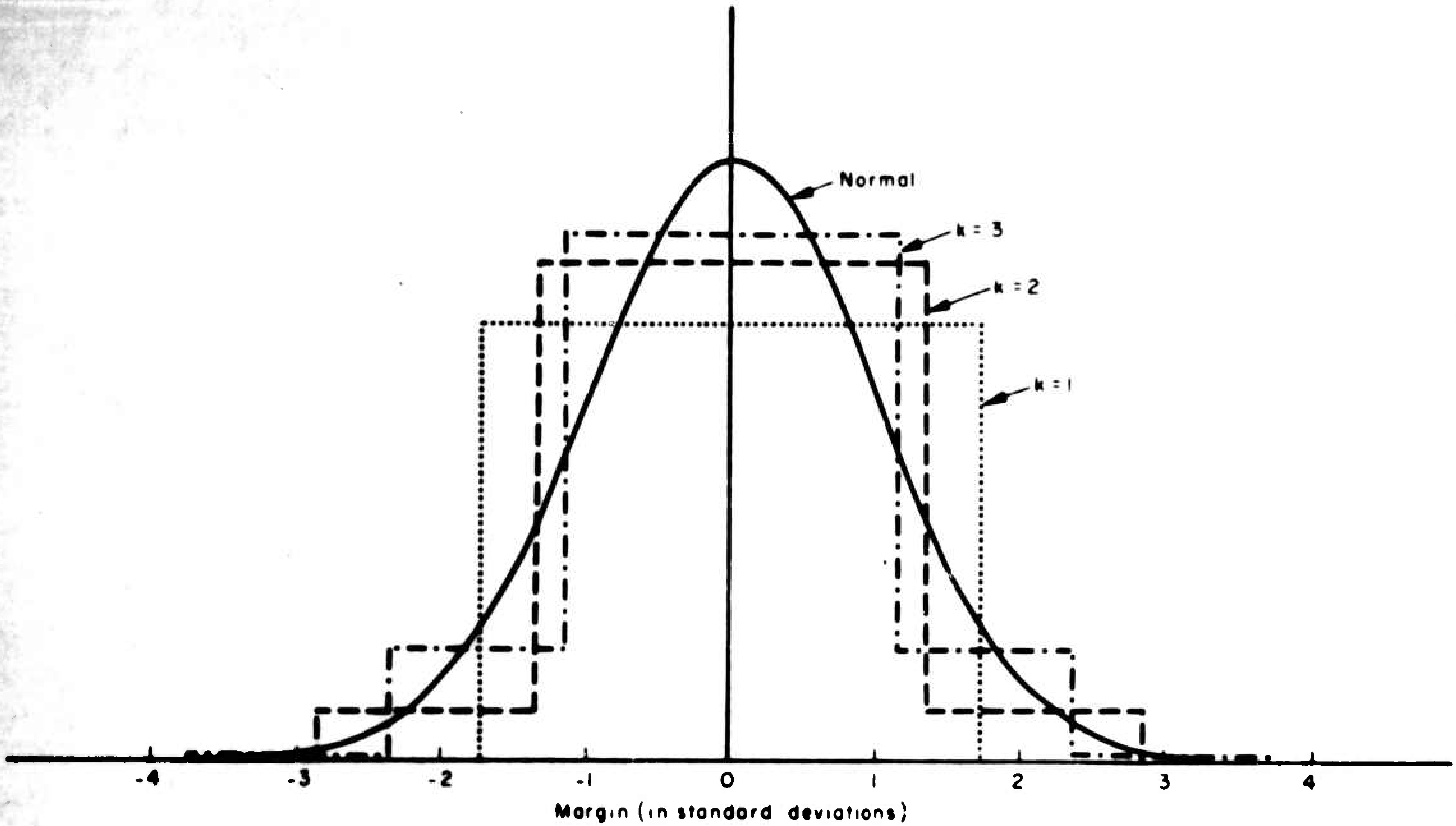


Fig. 1—Approximating a normal distribution by a mixture of uniform distributions: Frequency functions

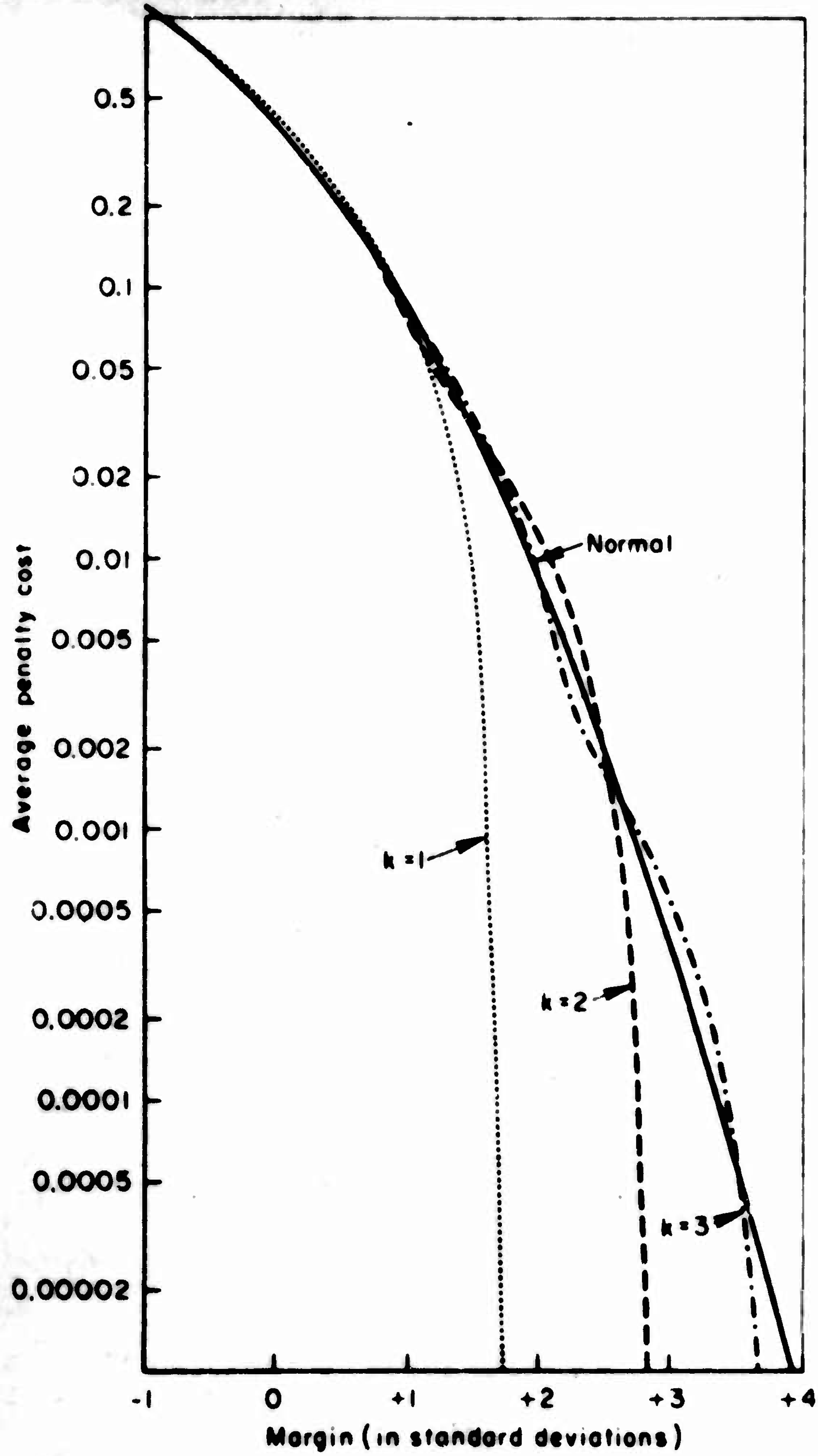


Fig. 2—Approximating a normal distribution by a mixture of uniform distributions: Average penalty costs ($\sigma = 1$, $f_i^- = 1$, $f_i^+ = 0$)

REFERENCES

Hurwicz, L. (1957), "Gradient Methods for Constrained Maxima," Operations Research, Vol. 5, pp. 258-265.

Beale, E. M. L. (1955), "On Minimizing a Convex Function Subject to Linear Inequalities," J.R. Statist. Soc. B, Vol. 17, pp. 173-184.

Beale, E. M. L. (1959), "On Quadratic Programming," Naval Research Logistics Quarterly, Vol. 6, pp. 227-243.

Charnes, A., W. W. Cooper and G. H. Symonds (1958), "Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil," Management Science, Vol. 4, pp. 235-263.

Charnes, A. and C. E. Lemke (1954), "Minimization of Nonlinear Separable Convex Functions," Naval Research Logistics Quarterly, Vol. 1, pp. 301-312.

Dantzig, G. B. (1955), "Linear Programming Under Uncertainty," Management Science, Vol. 1, pp. 197-206.

Dantzig, G. B. (1956), "Recent Advances in Linear Programming," Management Science, Vol. 2, pp. 131-144.

Dantzig, G. B. and A. Madansky (1960), "On the Solution of Two-Stage Linear Programs Under Uncertainty." Paper presented at the Proceedings of the Fourth Berkeley Symposium on Probability and Statistics, June 20-24, 1960.

Dantzig, G. B. and P. Wolfe (1960), "Decomposition Principle for Linear Programs," Operations Research, Vol. 8, pp. 101-111.

Elmaghraby, S. E. (1959), "An Approach to Linear Programming Under Uncertainty," Operations Research, Vol. 7, pp. 208-216.

Elmaghraby, S. E. (1960), "Allocation Under Uncertainty when the Demand has Continuous D. F.," Management Science, Vol. 6, pp. 270-294.

Ferguson, A. R. and G. B. Dantzig (1956), "The Allocation of Aircraft to Routes — An Example of Linear Programming under Uncertain Demand," Management Science, Vol. 3, pp. 45-73.

Madansky, A. (1960), "Methods of Solution of Linear Programs Under Uncertainty," The RAND Corporation paper, P-2132.